UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/732,809	12/11/2003	Johannes Heinecke	324-163	5874

7590   06/22/2009

LOWE HAUPTMAN GILMAN & BERNER, LLP
Suite 300
1700 Diagonal Road
Alexandria, VA 22314

EXAMINER
SHAH, PARAS D

ART UNIT	PAPER NUMBER
2626	

MAIL DATE	DELIVERY MODE
06/22/2009	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

PTOL-90A (Rev. 04/07)

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE _3_ MONTH(S) OR THIRTY (30) DAYS,
WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.
- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed
  after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
  Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any
  earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on _08 April 2009_.

2a)☐ This action is **FINAL**.   2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is
closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) _1,3-8,10 and 12_ is/are pending in the application.

    4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) _1,3,4, 6-8, and 12_ is/are rejected.

7)☒ Claim(s) _5 and 10_ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

    Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

    Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

    a)☐ All b)☐ Some * c)☐ None of:

      1.☐ Certified copies of the priority documents have been received.

      2.☐ Certified copies of the priority documents have been received in Application No. _____.

      3.☐ Copies of the certified copies of the priority documents have been received in this National Stage
application from the International Bureau (PCT Rule 17.2(a)).

    * See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1)☐ Notice of References Cited (PTO-892)

2)☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3)☐ Information Disclosure Statement(s) (PTO/SB/08)
    Paper No(s)/Mail Date _____.

4)☐ Interview Summary (PTO-413)
    Paper No(s)/Mail Date. _____.

5)☐ Notice of Informal Patent Application

6)☐ Other: _____.

DETAILED ACTION

1.   This communication is in response to applicant's arguments and amendments

filed on 04/08/2009. Claims 1-12 are currently pending in the application, with claim 12

being newly added and claims 2, 9, and 11 being cancelled. The Applicants'

amendment and remarks have been carefully considered, and are not persuasive.

2.   All previous objections and rejections directed to the Applicant's disclosure and

claims not discussed in this Office Action have been withdrawn by the Examiner.

Response to Amendments and Arguments

3.   Applicant's arguments see page 7-10 of applicant's remarks, filed 04/08/2009,

with respect to claims 1-12 have been fully considered but they are not persuasive.

As to claims 1-4 and 6-9, the Applicant argues that Van den Akker fails to teach

"constructing for each extracted word a plurality of character strings, including prefixes,

suffixes, and infixes with overlap. The Examiner respectfully disagrees with this

assertion. Van den Akker in col. 8, lines 5-12, describes the concept of bound

morphemes of which a single word may be composed of affixes. Further, in col. 8, lines

63-col. 9, lines 3, Van den Akker teaches that other word portions containing other

types of morphemes and further in col. 20, lines 36-43 states that a combination of word

portions can be extracted, constituting a plurality as recited in claim 1. Further, the

varying lengths for the extracted portions are found in col. 12, lines 27-32, where the

extracted suffix can be three characters or less when less then the predetermined

threshold. Hence, when viewed in light of the above teachings of using various other

types of word portions, extracting portions of varying sizes would have been obvious consistent with the teachings for the suffix extraction as described in the reference. The secondary reference of de Campos teaches the extraction of strings with overlap by the use of n-grams. In col. 10, lines 46-65, where a trigram window is shifted by one character. The use of such window allows overlapping sequences to occur.

Further, the Applicant argues that Van den Akker does not base the score of the character string based on the position of the first character string. The Examiner respectfully disagrees with this assertion. Van den Akker assigns a score to each extracted word portion based on the input text, where the word portion is based on "suffix type" extracted as in col. 9, lines 21-22. Further, a score for each word portion is determined based on frequency (see col.9, 35-41, probability value for each word portion). The position of the character string is taken into consideration by van den Akker since the position of the character string defines the word portion. The word portions identified are given a score, where the score is dependent on the suffix extracted. This extraction utilizes position information within the word for the identification (i.e. location of suffix in the word and the number of characters). Further, it should be noted that the claim recites the coefficient being dependent on the identification character string, where such limitation is broad enough to read on the extracted suffix based on its location in a word, where the last characters in a words are used to identify the suffix. The Applicant's further assert that the position and frequency of a character chain in a extracted word are distinctly identified. However, the independent claim does not recite any use of frequency information until claims 3 and 5.

Further, all claims except claims containing allowable subject matter (claims 5,

10, and 12) dependent upon the rejected base claim are rejected for similar reasons as

noted above.

Claim Rejections - 35 USC § 112

4.   The following is a quotation of the second paragraph of 35 U.S.C. 112:

> The specification shall conclude with one or more claims particularly pointing out and distinctly
> claiming the subject matter which the applicant regards as his invention.

5.   Claims 1, 8, and 12 are rejected under 35 U.S.C. 112, second paragraph, as

being indefinite for failing to particularly point out and distinctly claim the subject matter

which applicant regards as the invention. It is unclear as to what the applicant is

intending by "a first character string" and "a second character string" in the second to

last paragraph of the independent claims. For example, it is unclear as to whether these

first and second character strings are related to the prestored character strings or they

are from the extracted character strings. Further, the scoring for one language based

on the found character strings would occur only if the former interpretation occurs since

a matching between character strings of one language would be matched with the

extracted character string from the extracted word. Hence, for the purposes of compact

prosecution they were intended to mean first or second character strings from the

prestored first or second character strings, respectively. The Applicant is requested to

fully describe the score calculation as recited in the 2nd to last paragraph since the

scoring is unclear with respect to the first and second character strings, where such

character strings are either related to the prestored character strings or not. It is

suggested that the formula in Figure 1, step 9 be incorporated and each variable

defined.

Claim Rejections - 35 USC § 101

6.   35 U.S.C. 101 reads as follows:

> Whoever invents or discovers any new and useful process, machine, manufacture, or composition of
> matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the
> conditions and requirements of this title.

Claim 8 and 10 are rejected under 35 U.S.C. 101 as not falling within one of the

four statutory categories of invention.   Supreme Court precedent[1] and recent Federal

Circuit decisions[2] indicate that a statutory "process" under 35 U.S.C. 101 must (1) be

tied to another statutory category (such as a particular apparatus), or (2) transform

underlying subject matter (such as an article or material) to a different state or thing.

While the instant claim(s) recite a series of steps or acts to be performed, the claim(s)

neither transform underlying subject matter nor positively tie to another statutory

category that accomplishes the claimed method steps, and therefore do not qualify as a

statutory process For example the language identification method including steps

of prestoring, analyzing, and comparing is of sufficient breadth that it would be

reasonably interpreted as a series of steps completely performed mentally, verbally or

without a machine. The Applicant has provided no explicit and deliberate definitions of

"prestoring", "analyzing" or "comparing" to limit the steps to the language identification

being done by a machine," and the claim language itself is sufficiently broad to read on

---

[1] *Diamond v. Diehr*, 450 U.S. 175, 184 (1981); *Parker v. Flook*, 437 U.S. 584, 588 n.9 (1978); *Gottschalk v. Benson*, 409 U.S. 63, 70 (1972); *Cochrane v. Deener*, 94 U.S. 780, 787-88 (1876).
[2] *In re Bilski*, 88 USPQ2d 1385 (Fed. Cir. 2008).

a human having two separate pieces of paper containing a list of character sequences

for plural languages and another piece of paper with character strings that infrequently

occur. Then, taking a word from a word document from a computer and writing such

word on another paper by extracting various character sequences for that word. Using

information from the two separate pieces of paper and the character strings from the

word taken, making a comparison based on the position of the string in one language

and assigning a score based on number of characters found in one language and

adjusting the score based on the two lists and performing such analysis for each

language. Then, choosing the language that had the highest score as the identified

language.

Claim Rejections - 35 USC § 103

7.   The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all

obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set
> forth in section 102 of this title, if the differences between the subject matter sought to be patented and
> the prior art are such that the subject matter as a whole would have been obvious at the time the
> invention was made to a person having ordinary skill in the art to which said subject matter pertains.
> Patentability shall not be negatived by the manner in which the invention was made.

8.   **Claims 1-4, 6-9** are rejected under 35 U.S.C. 103(a) as being unpatentable over

VAN DEN AKKER (Patent No.: US 6,415,250) in view of DE CAMPOS (Patent No.: US

6,272,456).

9.   Regarding **claims 1 and 8**, VAN DEN AKKER teaches a device for automatically

identifying the language of a digital text ("automatic language identification system",

column 6, line 40), comprising:

means for prestoring (see col. 11, lines 3-7, memory 20, and 30 and see col. 6,

line 56-61, where the storage and memory devices used in conjunction with the system)

first character strings, including prefixes, suffixes, and infixes (see Figure 5A, and col. 8,

lines 5-12, describes the concept of bound morphemes of which a single word may be

composed of affixes and in col. 8, lines 63-col. 9, lines 3, other word portions containing

other types of morphemes) of different lengths from words of a plurality of

predetermined languages (see Figure 5A, where suffixes for plural language are shown

of varying lengths), that occur frequently anywhere respectively in said words of said

plurality of predetermined languages ("probability table 304 includes an entry for every

selected word portion 303 that occurs in at least one of the language corpuses 309",

column 10, lines 18-20);

means for prestoring second character strings of different lengths that are

atypical anywhere respectively in words of said predetermined languages ("probability

table 304 includes an entry for every selected word portion 303 that occurs in at least

one of the language corpuses 309", column 10, lines 18-20 and see col. 9, lines 1-16,

where variety of corpora are used.);

means for analyzing words (see col. 7,line 50-58, where software implemented

on a computer system is used for input and identification and see col. 9, lines 6-7,

language corpus analyzer)) extracted from said digital text thereby constructing for each

extracted word a plurality of character strings (see col. 20, lines 36-43 states that a

combination of word portions can be extracted, constituting a plurality) contained in said

extracted word ("word portions extracted from the input text 301", column 10, lines 39-

40), including prefixes, suffixes, and infixes, (see col. 8, lines 9-12 and col. 8, lines 63-

col. 9, lines 3, Van den Akker teaches that other word portions containing other types of

morphemes and further in col. 20, lines 36-43 states that a combination of word portions

can be extracted such as prefix and suffix) and different lengths (see col. 12, lines 26-

32, when a suffix is extracted based on varying lengths) lying between one character

and the number of characters in said extracted word ("more or less characters may be

included in the predetermined number of characters", column 9, lines 22-23);

means for comparing (see col. 7,line 50-58, where software implemented on a

computer system and see col. 9, lines 6-7, language corpus analyzer) each of said

plurality of character strings (see col. 20, lines 36-43 states that a combination of word

portions can be extracted, constituting a plurality) contained in each said extracted

word individually to said first and second prestored character strings of each

predetermined language so that whenever a first character string is found in said

extracted word a score associated with said one language is increased by a first

coefficient depending on the position of said first character string of said one determined

language found in said extracted word (see column 10, lines 37-42, and FIG. 6, the

suffixes are used for scoring, meaning the values are dependent on the position of the

characters, since characters from the suffix are used and Figure 7, where the unknown

text 301 is input and compared using language determiner 706, col. 15, lines 7-15,

where the score is determined (i.e. probability) for language) and whenever a second

character string is found in said extracted word a respective second coefficient that is

associated with said found second character string (see FIG. 6, "probability table 304 is

altered to include predetermined negative values for those word portions which do not

appear in a language corpus 309", column 13, lines 62-64) (e.g. The reference shows

the comparison of an extracted word to multiple language corpus, which is seen in

Figures 6 and 7. hence, corresponding probabilities are increases or decreased based

on probable occurrences of the string); and

means for comparing (see col. 10, lines 33-35, language identification engine)

said scores for said text associated with said predetermined languages in order to

determine the highest of said scores, which identifies the language of said text ("the

largest accumulated relative likelihood value, provided it exceeds zero, identifies the

language of the input text 301", column 10, lines 42-44).

However, VAN DEN AKKER does not disclose that whenever a second character

string is found in said extracted word in said extracted word, said score is decreased by

a respective second coefficient and said respective second character coefficient

increasing as the probability of said found character string in said each determined

language decreases.

In the same field of language identification, DE CAMPOS teaches the use of

character string extraction from a words with overlap (see col. 10, lines 46-65, where a

trigram window is shifted by one character) and whenever a second character string is

found in said extracted word in said extracted word (see col. 3, lines 60-67, if the

character string is found in many languages, therefore a second character string is

analyzed), said score is decreased by a respective second coefficient (see col. 3, lines

65-66, score is decreased if found in many languages) and said respective second

coefficient increasing as the probability of said found character string in said one

determined language decreases (see col. 3, lines 60-67, score is increased for

infrequently appearing strings for the specific language is increased, but if it occurs in

another languages score decreases. Thus, the increase in score only occurs if the

match occurs in few languages, where the other languages do not contain such term

and collectively the probability of such word in all languages decreases. Although a

second coefficient is not used it would have been obvious to one skilled in the art to add

two separate coefficients rather than increasing or decreasing for the objective of

discriminating between infrequent sequences (i.e. score (language 1) =alpha – beta)

(see DE CAMPOS, col. 4, lines 62-65)) (e.g. Further, the claimed limitation of the

coefficient increasing is evident by the decrease for frequently occurring words in other

languages, which entails that a decreasing score lead to a lesser determination that the

extracted word came from that language).

Therefore, it would have been obvious to a person of ordinary skill in the art at

the time the invention was made to use the coefficient modification of DE CAMPOS in

the language identification system of VAN DEN AKKER in order to discriminate

languages in identifying languages with infrequently appearing sequence (see DE

CAMPOS, col. 3, lines 67-col. 4 lines 1-4 and lines 62-65).

As to claim 9, van den Akker teaches all of the limitations as in claims 1 and 8 above, and further teaches the computer readable storage device (see col. 6, lines 56-57) storing software in conjunction with a processor (see col. 6, lines 59 and lines 42).

10.   Regarding **claim 3**, VAN DEN AKKER in view of DE CAMPOS teach all of the limitations as in claim 1 above. VAN DEN AKKER further teaches that said first coefficient of a first character string in said extracted word depends on the frequency of said character string in said determined language ("frequency value indicative of the number of times the selected word portion was found within the corresponding language corpus 309", column 9, lines 36-38).

11.   Regarding **claim 4**, VAN DEN AKKER in view of DE CAMPOS teach all of the limitations as in claim 1 above. DE CAMPOS further teaches that said first coefficient of a first character string in said extracted word depends on the length of said character string ("the language ID program module 36 is looking for the longest match to the test letter sequence of letters appearing in the window", column 13, lines 54-56).

12.   Regarding **claim 6**, VAN DEN AKKER in view of DE CAMPOS teach all of the limitations as in claim 1 above. VAN DEN AKKER further teaches comparator means for comparing each of said extracted words from said text with frequent words in said determined language and initially listed in storage means (see col. 11, lines 3-7, memory 20, and 30 and see col. 6, line 56-61, where the storage and memory devices

used in conjunction with the system) so that whenever a frequent word is found in said

text said score for said determined language is increased only by a coefficient

depending on the frequency of said extracted word in said determined language

("identification engine 306 searches the probability table 304 for each of the

morphologically-significant word portions extracted from the input text 301, summing the

relative probability values associated with each language for each of the extracted word

portions", column 10, lines 37-42) (e.g. Depending on whether word portion is found the

probability values are summed increasing the score).

Furthermore, DE CAMPOS teaches increasing the score for one of the

languages when the longest match is found in a few languages.

13.   Regarding **claim 7**, VAN DEN AKKER in view of DE CAMPOS teach all of the

limitations as in claim 1 above. VAN DEN AKKER further teaches the storage means.

(see col. 11, lines 3-7, memory 20, and 30 and see col. 6, line 56-61, where the storage

and memory devices used in conjunction with the system).

DE CAMPOS further teaches comparator means for comparing each of said

extracted words from said text with frequent words in said determined language and

initially listed in storage means so that whenever a frequent word is found in said text

said score for said determined language is increased only by a coefficient depending on

the length of said frequent word ("the language ID program module 36 is looking for the

longest match to the test letter sequence of letters appearing in the window", column 13,

lines 54-56 and col. 18, lines 26-31, based on length of a word the longer matches are increased in terms of score value).

Allowable Subject Matter

14.   Claim 12 would be allowable if rewritten or amended to overcome the rejection(s) under 35 U.S.C. 112, 2nd paragraph, set forth in this Office action.

15.   Claims 5 and 10 are objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.

16.   The following is a statement of reasons for the indication of allowable subject matter: DE CAMPOS teaches a score for each language based upon a frequency parameter in the n-gram profiles corresponding to the length of the longest match. VAN DEN AKKER teaches a probability value corresponds directly to the frequency FR. However, none of the prior art references or in combination thereof teach the coefficient of a first character string equal to PO(FR + LON), as recited in claim 5.

Conclusion

17.   Any inquiry concerning this communication or earlier communications from the examiner should be directed to PARAS SHAH whose telephone number is (571)270-1650. The examiner can normally be reached on MON.-THURS. 7:00a.m.-4:00p.m. EST.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's

supervisor, David Hudspeth can be reached on (571)272-7843. The fax phone number

for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the

Patent Application Information Retrieval (PAIR) system. Status information for

published applications may be obtained from either Private PAIR or Public PAIR.

Status information for unpublished applications is available through Private PAIR only.

For more information about the PAIR system, see http://pair-direct.uspto.gov. Should

you have questions on access to the Private PAIR system, contact the Electronic

Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a

USPTO Customer Service Representative or access to the automated information

system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/David R Hudspeth/
Supervisory Patent Examiner, Art Unit 2626

/P. S./
Examiner, Art Unit 2626

06/17/2009